

# Data Processing Methods

---

## Primary Data Collection

The California Public Utilities Commission (CPUC) sent out a Data Request to broadband providers to initiate the Round 4 data collection. Potential providers were widely encouraged to submit broadband service availability data to the CPUC. We expressed our preference for providers to use a file geodatabase format when possible while tabular data was also accepted. Data submittal instructions were posted to assist providers with the process, template files, sample shapefiles and record formats on the CPUC Broadband Mapping Website at:

<http://www.cpuc.ca.gov/PUC/Telco/Information+for+providing+service/BroadBand+Mapping.htm>

The data submittal instructions guide each provider to wireless and/or wireline datasets which are separated into sections for those with GIS data (shapefiles/FGDB) and those without GIS data (text/excel files). For providers with GIS capabilities, statewide census block and TIGER/Line shapefiles were provided on the CPUC website for download and use for their data submission. The square mileage of each block was calculated in advance in the sample census block shapefile. Providers could then easily determine which blocks in their footprint were less than two square miles and which were two square miles or greater and therefore needed to be represented using the road segment shapefile. For providers without GIS capabilities, excel spreadsheets were provided incorporating record field formats adhering to the NOFA data submittal requirements.

## Community Anchor Institutions (CAI)

CAI data initially came from the eligible entries of California Teleconnect Fund (CTF) program. The CTF program provides 50% discounts on telecommunications bills for qualifying schools, libraries, government-owned and operated hospitals and health clinics, and other community based organizations, thus providing a good initial list of CAIs. The CAI addresses were geocoded to point locations and loaded into the file geodatabase. Technology of transmission and speeds data were included and identified either through information received from the Institutions themselves (as in the case of libraries) or from those service providers who responded to our request for such information. To provide CAI ID information (as in the case of schools), the California Department of Education search engine website (<http://www.cde.ca.gov/re/sd/>) was utilized.

## Provider Participation

A total of 71 providers participated in Round 4 data collection. These providers comprise over 99.9% of the total broadband connections in California reported to the FCC on form 477, which constitutes a very complete set of data

## **CPUC Initial Data Verification**

After obtaining files submitted by the providers, a data inventory spreadsheet was used to reflect the assigned GIS team member and record count. Each file was reviewed against the GIS data model posted in the SBDD Network website to see if mandatory fields were filled in, and each field was checked for the appropriate range of values. Where possible, team members loaded the submitted data into the corresponding geodatabase table to make certain that appropriate field headers were used and that each field contained the correct data type. When data was found to be missing or incorrect, the provider was contacted and the issue was documented in a separate provider spreadsheet. Some providers submitted nearly perfect data sets while others gave incomplete, unexpected, or incorrect data. New information, correspondence with the providers, and fixes made by the CPUC were also documented in each provider spreadsheet.

## **Chico GIC Geoprocessing**

After the initial CPUC review, data was transferred to the Geographical Information Center (GIC) at CSU Chico for geocoding, geomatching, propagation of wireless service by antenna, and validation of geographic data. In those cases where the CPUC had received street address level data from broadband providers, such addresses were assigned a point location, (geocoded) and then geomatched to census blocks and street segments.

Providers who offer wireless service but could not submit a shapefile or geographic representation of their service area gave tabular antenna information. Wireless antenna parameters were used to model a service area and shapefiles were created for each provider. The wireless propagation model is based on the Longley-Rice, Irregular Terrain propagation model. Individual unit specifications are used to measure performance based on frequency, transmit power, receiver sensitivity, antenna gain, and height. Signal coverage patterns are produced for each individual unit taking into account terrain and vegetation features that may hinder signal dispersion.

## **CPUC Final Data Verification**

The resulting datasets were delivered from Chico to the CPUC in the SBDD transfer model geodatabase for final review and verification. Data sets were checked again and reviewed for unexpected changes resulting from the geocoding /geomatching process. Geoprocessed data was visually reviewed using ArcGIS to verify service area footprints, and the SBDD check submission Python script was run on each dataset to identify unexpected values.

## **Deliverable Data**

The final dataset is delivered to the NTIA/FCC in file geodatabase format with the following feature classes:

BB\_ConnectionPoint\_LastMile – not required per Clarification to the NOFA.

BB\_ConnectionPoint\_MiddleMile – Point between the local “last mile” network and the middle mile network which goes on to connect to the internet backbone. This is a confidential dataset.

BB\_Service\_Address – not included per the CPUC NDA.

BB\_Service\_CAIstitutions – Community Anchor Institutions: points geocoded from address lists

BB\_Service\_CensusBlock – Broadband availability polygons for areas less than 2 square miles

BB\_Service\_Overview – Service overview by County including Subscriber Weighted Nominal Speed

BB\_Service\_RoadSegment – Broadband availability line segments for areas 2 square miles and greater

BB\_Service\_Wireless – Wireless service area polygons.

## **Planned Validation Methods**

The following validation methods will be conducted on Round 4 data. Detailed maps showing submitted service area footprints and areas that could not be validated will be distributed to each provider for feedback.

### **FCC Form 477**

FCC Form 477 collects information about broadband connections to end user locations, wired and wireless local telephone services, and interconnected Voice over Internet Protocol (VoIP) services, in individual states, at the Census Tract level. A shape file was created for each provider reflecting the presumed availability of broadband service at each census tract where the provider reported customers to their fixed broadband service. These layers were used to cross reference ISP data submissions to the CPUC.

### **ID Insight, BroadBand Scout**

BroadBand Scout is a third party comprehensive and unbiased data specifically designed to show the carriers, connectivity, speed and usage details of the national broadband landscape. ID Insight’s patent-pending process analyzes hundreds of millions of internet transactions that link a consumer’s physical address to their internet carrier. BroadBand Scout data is provided as tabular point locations and geomatched to the census block level where less the two square miles in area and to the street segment level where census blocks are greater than two square miles in area. A shape file was created for each provider reflecting the presumed availability of broadband service at each census block or street segment where Broadband Scout reported online customer transactions. These layers were used to cross reference ISP data submissions to the CPUC.

### **TeleAtlas Wire Center**

The Wire Center Premium product is a comprehensive database for mapping and analyzing wire center service areas. It forms the backbone of the Tele Atlas® Telecommunication Products line. This product lists every Local Exchange Carrier (LEC) landline wire center in the United States. The term “wire center” refers to the location where the telephone company terminates the local lines; this is usually the same location as a central office, although a wire center might house one or more central offices. Buffers were

created at 12,000 feet and 18,000 feet from provided Wire Center point datasets to cross reference ISP data submissions to the CPUC.

### TeleAtlas Wire Center Region

The Wire Center Premium product is a comprehensive database for mapping and analyzing wire center service areas. It forms the backbone of the Tele Atlas® Telecommunication Products line. This product lists every Local Exchange Carrier (LEC) landline wire center in the United States. The wire center boundary is a representation of the area served by all of the switching equipment housed at that physical location. Wire Center Region polygon GIS layers were provided and used for cross referencing ISP data submissions to the CPUC.

### FCC Consumer Broadband Test (Non-Mobile App)

The FCC offers an Online Consumer Broadband Test. FCC's Online Consumer Broadband Test collects information regarding the location of the client, the engine used to provide the speed test, download speed, upload speed, latency, jitter, packet loss, minimum round trip time, maximum round trip time, and average round trip time at a specified point location. A shape file was created to represent each location at which speed tests were performed based on geocoded address records. All point locations were then geomatched to the census block level where less the two square miles in area and to street segment level where census blocks are greater than two square miles in area. These layers were used to cross reference ISP data submissions to the CPUC where sub-broadband speeds were reported and/or where there were no tests performed.

### FCC Consumer Broadband Test (Mobile App)

The FCC offers a Mobile Consumer Broadband Test for the Apple iPhone and Android mobile platforms. The official name of the App is the **FCC Broadband Test**. This tool can be downloaded to an Apple or Android enabled device by accessing the App Store or App Market on a handheld phone. FCC's Mobile Consumer Broadband Test collects information regarding the location of the client, the client's operating system, the engine used to provide the speed test (always Ookla for mobile tests), download speed, upload speed, and latency, at a specified point location. A shape file was created to represent each location at which speed tests were performed based on latitude and longitude coordinate pairs. All point locations were then geoprocessed to the census block level where less the two square miles in area and to street segment level where census blocks are greater than two square miles in area. These layers were used to cross reference ISP data submissions to the CPUC where sub-broadband speeds were reported and/or where there were no tests performed.

### FCC Broadband Dead Zone Reporting Form

The FCC offers a Broadband Dead Zone Reporting Form for recording address or city level queries against the National Broadband Map, that failed to return any providers at the specified location, or alternately, where a user may know that no service is provided at a specific address. FCC's Broadband Dead Zone Form collects information regarding the location of the client, whether the client has internet access at their home, what type of internet access the client has at their home, and whether or not the client would be interested in purchasing broadband internet if service options were available. A shape file was created to represent each location for which dead zone forms were filled out based on

geocoded address records. All point locations were then geomatched to the census block level, where less than two square miles in area, and to street segment level, where census blocks are greater than two square miles in area. These layers were then used to cross reference ISP data submissions to the CPUC where dead zones and/or no services provided were reported.

### **California State Map Broadband Service Survey Feedback**

The CPUC offers the Broadband Service Survey within its interactive map. The survey records user feedback based on address, city, or zip code level queries against the State's Broadband Availability. It collects information regarding the location of the client, whether the client is accessing the internet from their home, place of business, or any other location, whether or not the client purchases broadband service, and if not, why they choose not to purchase broadband service. A shape file based on geocoded address records was created to represent each location for which service surveys were submitted where the respondent indicated non-subscription because of no broadband availability. All such point locations were then geomatched to the census block level, where less than two square miles in area and to the street segment level, where census blocks are greater than two square miles in area. These layers were then used to cross reference ISP data submissions to the CPUC

### **Chico GIC Data Validation Processes**

Each individual provider's submitted and/or created data was validated independently to all applicable validation methods. The following fields were added to each individual provider's data tables, where appropriate; FCC\_477 (FCC Form 477), BBSCOUT (ID Insight BroadBand Scout), TA\_WC\_REG (TeleAtlas Wire Center Region), WC\_VAL\_12K (TeleAtlas Wire Center 12,000 foot buffer), WC\_VAL\_18K (TeleAtlas Wire Center 18,000 foot buffer), VAL12k\_18k (TeleAtlas Wire Center 12,000 to 18,000 foot buffer ring), DEGRAD\_FT (TeleAtlas Wire Center distance), (FCC\_TST) FCC Consumer Broadband Test Non-Mobile App, (FCC\_MOBL) FCC Consumer Broadband Test Mobile App, (FCC\_DZ) FCC Broadband Dead Zone Reporting Form, and (CA\_SRVY) State Map Broadband Service Survey Feedback to record validation results and to allow symbology of discrepancies based on validation methods for further interaction with each provider to refine their data submissions. The final step was a summary statistics report of all validation results for all submitted providers. Summary statistics include validity counts and percentages for all validation methods, specific to provider and technology.

### **Wireline Census Block Validation**

A spatial selection was performed on Census Block data, either submitted by provider, or created from submitted address records through a geocoding/spatial selection process, to derive only those blocks which intersect polygons in a given validation layer. Counts are recorded as number of unique blocks which share geographic area with any given validation layer, compared to the total number of unique blocks submitted by, or created for, a given provider. Percentages are recorded as percentage of the total number of unique blocks which share geographic area with any given validation layer, compared to the total number of unique blocks submitted by, or created for, a given provider.

### **Wireline Street Segment Validation**

A spatial selection was performed on Street Segment data, either submitted by provider, or created from submitted address records through a geocoding/spatial selection process, to derive only those

segments which intersect polygons in a given validation layer. Counts are recorded as number of unique segments which share geographic area with any given validation layer, compared to the total number of unique segments submitted by, or created for, a given provider. Percentages are recorded as percentage of the total number of unique blocks which share geographic area with any given validation layer, compared to the total number of unique segments submitted by, or created for, a given provider.

### **Wireless Validation**

A spatial selection was performed on Wireless Availability data, either submitted by provider, or created from antenna location and specification information, to select only those polygons which intersect a given validation layer. Results are recorded as a percentage of the total geographic area of wireless coverage sharing geographic area with any given validation layer, compared to the total coverage area submitted by, or created for, a given provider.